DOCUMENT RESUME

ED 235 226 TM 830 647

AUTHOR Chafin, Ann E.
TITLE Setting a Standard for Standard Setting.
PUB DATE Apr 83
NOTE 18p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Basic Skills; Criterion Referenced Tests; *Cutting Scores; *Graduation Requirements; High Schools; *Minimum Competency Testing; *Quantitative Tests; State Departments of Education; *State Standards
IDENTIFIERS *Maryland Functional Mathematics Test; Maryland State Department of Education; *Standard Setting

ABSTRACT

The problem of setting standards has always been an issue in the educational community. This paper offers an outline of one state's procedure for setting a passing score for a competency-based test in mathematics. The Maryland Functional Mathematics Test (MFMT) becomes a graduation prerequisite for the graduating class of 1987. The procedures used for gathering information for standard setting are very similar to the procedures developed by Richard Jaeger. The standard setting procedures were designed to meet the two key criteria of maximum participation and multiple types of information. This political-judgmental approach allowed teachers, citizens, administrators, students and parents in the State of Maryland to offer a recommendation on the passing score. Sets of recommended performance standards for the MFMT were secured from 4,729 individuals with only slight differences among the standards recommended. The summaries of their recommendations formed the basis for the decision making by the State Superintendent of Schools. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235226

SETTING A STANDARD FOR

STANDARD SETTING

ANN E. CHAFIN
MARYLAND STATE DEPARTMENT OF EDUCATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. E. Chafin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 617

A symposium presentation at the annual meeting of the National Council on Measurement in Education, Montreal, April 11-15, 1983.

SETTING A STANDARD FOR STANDARD SETTING

Abstract

The problem of setting standards has always been an issue in the educational community. This paper offers an outline of one state's procedure for setting a passing score for a competency-based test in mathematics. The Maryland Functional Mathematics Test (MFMT) becomes a graduation prerequisite for the graduating class of 1987.

The procedures used for gathering information for standard setting are very similar to the procedures developed by Dr. Richard Jaeger. This political-judgemental approach allowed teachers, citizens, administrators, students and parents in the State of Maryland to offer a recommendation on the passing score. The summaries of their recommendations formed the basis for the decision making by the State Superintendent of Schools.

SETTING A STANDARD FOR STANDARD SETTING

The problem of setting standards has always been an issue in the educational community. With more states and school districts requiring competency tests for certification, a number of researchers have become interested in developing a technology for standard setting. Many pitfalls in the process of setting standards are discussed in Gene Glass' well known "Standards and Criteria" (1978). Although thought provoking, this article offers little advice to those educators who must set standards. Both Hambleton (1980) and Popham (1981) have put forth very convincing arguments that standard setting is arbitrary but need not be capricious. Livingston and Zieky, in their pamphlet Passing Scores (1982) give a useful practitioner's guide to standard setting, again emphasizing that judgments are always involved but can be considered and thoughtful. Jaeger (1981) developed a political-judgmental approach to standard setting that was premised on the need for affected audiences to offer judgments. These judges were given different kinds of information to inform their choice of standards.

All of these articles and the thinking they represented formed the information base for the Maryland State Department of Education to begin the development of a standard setting procedure for various components of the Maryland Functional Testing Program.

MARYLAND FUNCTIONAL TESTING PROGRAM

Maryland uses a competency-based instructional program which includes criterion-referenced tests in the areas of functional reading, mathematics and writing. To insure quality instruction in the basics, the Maryland State Board of Education (MSBE) requires all students to pass these tests before they are eligible for high school graduation. Choosing a passing score for each content area becomes a sensitive and demanding issue. The Maryland Functional Reading Test (MFRT) has a standard already in existence. Subsequent forms are equated back to this level of performance. Standard setting for the Functional Writing Test is still under discussion since field testing of two writing prompts has just been completed. This paper will discuss the setting of a passing standard for the Maryland Functional Mathematics Test (MFMT) which was given under no-fault conditions in the Fall of 1982. However, because the standard that is set on the MFMT can be no more defensible than the assessment instrument upon which it is based, it is appropriate to precede the description of standard-setting with a brief history of the development of the MFMT.

MARYLAND FUNCTIONAL MATHEMATICS TEST

Development of the mathematics test began in 1978, with the identification of content areas that Maryland educators, parents and citizens felt were appropriate for coverage by a functional mathematics program. These content areas were adopted by the Maryland State Board of Education and subsequently disseminated throughout the state in the MSDE publication, Declared Competencies Index (1980).

The second step in the test development process took place in March of 1981 when a group of Maryland mathematics teachers and curriculum specialists used the 30 objectives in the Declared Competencies Index to develop test specifications for an assessment system. Seven different assessment domains were identified and the rules for generating test items to measure each domain carefully delineated. The resulting document, "Test Specifications for the Maryland Functional Mathematics Test," Maryland State Department of Education, Baltimore, November 1981, was used to guide the production of 546 mathematics test items, 78 items per assessment domain. One-half of these items were developed for use in the upper level version of the MFMT. This upper level form of the measure is given for the first time in the ninth grade and is used for graduation certification. The other half of the test items were developed for use in the lower level version of the MFMT, a seventh-grade examination that serves a diagnostic and instructional function.

In the Fall of 1981, all of the mathematics test items were field tested by representative samples of Maryland students. A total of 24,000 students took part in this field test, including 9,000 ninth-graders and 3,000 eleventh-graders who also participated in the tryout of the upper level MFMT items. Data analyses from this field test provided various indices of item quality which were subsequently used to guide the construction of the first operational forms of the MFMT. The field test results also yielded student performance data. With an operational test form in place, MSDE now must present information that will help guide the State Superintendent of Schools as he makes this important decision regarding an appropriate passing standard.

CRITERIA FOR STANDARD-SETTING

Establishing the passing standard for a certification examination can be the most vulnerable operation in the entire development of that instrument. Since the introduction and widespread use of minimum competency tests, the issue of cut-score determination has received a great deal of attention. Two key issues seem to emerge in all such considerations. First, standard-setting involves political as well as psychometric concerns. To be credible, a standard-setting operation should involve as many different concerned constituencies as is practically possible. The more individuals who participate in the process of recommending a cut-score, the more credibility that process will possess.

Second, standard-setting is at heart a judgmental operation. Although there are many different procedures for standard-setting currently in use, all of them rely at some point on human judgment. Judgmentally-based decisions, if not made carefully, will always be subject to the criticism that they are arbitrary and inconsistent. One strategy for avoiding such charges is to provide the individual or group who must make the final decision regarding a cut-score with as much information as possible. Multiple sources of information will, hopefully, allow for the most informed judgment possible.

The standard-setting procedures used for the Maryland Functional Mathematics Tests were designed to meet the two key criteria of maximum participation and multiple types of information. The question "How well must a student do in order to pass the Maryland Functional Mathematics Tests?" will ultimately be answered by the State Superintendent of Schools. The decision-maker will be provided with information of three different types; namely, test item-based judgments, test specifications-based judgments, and student performance data. The following pages elaborate on the types of information collected and the results.

ITEM-BASED JUDGMENTS

On Saturday, June 12, 1982, at three separate sites in Maryland, 61 individuals reviewed the Maryland Functional Mathematics Test (MFMT), and then offered recommendations regarding an appropriate passing standard for that test.

Three categories of standard advisors were represented at the workshops, namely, citizens, mathematics teachers, and school administrators (including counselors). In all, there were 28 citizens, 10 mathematics teachers, and 10 administrators who served as standard advisors. The three sites were Annapolis (19 participants), Easton (28 participants), and Frederick (14 participants). A substantially larger number of individuals had been invited to participate in the workshops via invitations directed to Maryland's 24 districts.

Recommendation One

The first recommendations secured at each site were based on individuals' taking an actual form of the test, then answering Yes or No, on an item-by-item basis, the following question: "Should every high school graduate in Maryland be able to answer this item correctly?" Of the 78 items rated, each advisor's total of Yes ratings (that is, the items which the advisor thought should be answered correctly by every high school graduate) constituted a recommended passing standard based on a review of the test items. Results of these initial recommendations are presented in Table 1.

TABLE 1

Initial Recommendations: Suggested Passing Standards Based on a Review of the Test Items.

Citizens (n=35)		Teachers (n=14)		Administrators (n=12)	
Items	Percent	Items	Percent	Items	Percent
69.9	89%	68.9	88%	69.7	89%

Recommendation Two

After the initial recommendations were secured, participants were provided with a set of field test performance data based on a Fall 1981 tryout of the test items with approximately 9,000 ninth-graders and 3,000 eleventh-graders. The standard advisors were provided with average performances and item-by-item difficulties. The actual performance data sheets supplied are provided in the Appendix. Again, participants answered the same Yes/No question on an item-by-item basis, but were encouraged to make their ratings in light of the performance data with which they had been supplied. Totals of Yes ratings were obtained for all participants. Each advisor's total number of Yes ratings can be thought of as a recommended passing standard based on a review of the test items plus student performance data. Results of the second set of recommendations are presented in Table 2.

TABLE 2

Second Recommendations: Suggested Passing Standards Based on a Review of the Test Items Plus Student Performance Data.

Citizens (n-35)		Teachers (n=14)		Administrators (n=12)	
Items	Percent	Items	Percent	Items	Percent
68.5	88%	66.2	85%	69.7	89%

Recommendation Three

The data from the second ratings were relayed by telephone from each workshop to a central site, aggregated, then telephoned back to each workshop. Participants were given peer feedback in the form of the statewide average recommended standard for their type of standard advisors(i.e., citizens, mathematics teachers, or administrators). Once more, participants were asked for a recommendation, but this time they were asked for only one response, namely, the "total number of items a student must answer correctly to pass." This third round of judgments led to a recommended passing standard based on a review of the test items, plus student performance data, plus peer feedback. Results of the third set of recommendations are presented in Table 3.

TABLE 3

Third Recommendations: Suggested Passing Standards Based on a Review of the Test Items, Plus Student Performance Data, Plus Peer Feedback.

Citizens (n=28)		Teachers (n=10)		Administrators (n=10)	
Items	Percent	Items	Percent	Items	Percent
66.6	85%	61.2	78%	65.3	84%

It should be noted that the final rating called for a single suggested passing level rather than a total of item-by-item Yes/No responses. This might account for the reduced passing standards seen from first to third ratings in all three categories of standard advisors.

In summary, a small group of Maryland citizens, teachers, and administrators, having become familiar with the Maryland Functional Mathematics Test by actually completing it, supplied three different sets of recommendations regarding a suitable passing standard for the test.

SPECIFICATIONS-BASED JUDGMENTS

During the third week of May 1982, packets of materials were mailed to principals of all Maryland secondary schools. In all, 312 sets of materials were mailed. The covering letter, authored by a high ranking official in the Maryland State Department of Education, described the purpose of the standard-advising survey, invited the principal's voluntary participation, provided directions for distributing the enclosed data-gathering instruments, and indicated how the school's report form was to be returned.

Briefly, participating schools were to distribute materials to approximately 10 teachers, 25 parents of twelfth-grade students and 25 twelfth-grade students. The summaries were to be returned in stamped, self-addressed envelopes by June 7.

By June 25, responses were considered final. Materials were received from 124 secondary schools, or, 40 percent of the schools originally invited to participate in this phase of the standard-setting process. Recalling, however, that at each school the principal was asked to seek the participation of numerous individuals, the total numbers of teachers, parents, and students who recommended a standard for the MFMT were substantially larger than the 124 schools from which responses were secured. In all, 1,255 educators (teachers), 1,268

parents, and 2,206 students supplied preferences regarding an appropriate standard for the MFMT. A single response form from each school was supplied indicating the mean recommended passing standard (average percent correct score) for each of the three groups (educators, parents, and students). The number of individuals per group was also supplied from each school. Thus, each school's mean data became the unit of analysis for this study.

Each individual receiving a set of materials was provided with information describing the MFMT, as well as sample items from each of the test's seven content domains. In addition, one half of the schools polled were given materials containing a single sheet summarizing the results of a Fall, 1981 field test of MFMT items. The purpose of this variation in the materials was to discern whether such performance data would appreciably influence the standards recommended by participants. A copy of the performance data sheet is included in the Appendix.

The results of the survey are presented in Tables 1, 2, and 3. In Table 1, the reader will find the recommendations of the educators, parents, and students from the 58 schools receiving materials without the performance data sheet. In Table 2, the same data are presented for individuals from the 66 schools receiving materials which contained the performance data sheet. In Table 3, results of both groups are combined for the 124 schools. Remember, the number of individuals reported reflects the total number of persons contributing to the school averages.

Table 1. Means and Standard Deviations of Passing Standards Recommended by Educators, Parents, and Students from 58 Schools Receiving Materials Without Field-Test Performance Data.

Educators			Parents			Students		
n	Mean %Correct Recommended	S.D.	n	Mean %Correct Recommended	S.D.	n	Mean %Correct Recommended	S.D.
561*	77.8	4.5	590*	77.8	3.9	1,012*	76.9	4.4

*Numbers of individuals contributing to school averages.

Table 2. Means and Standard Deviations of Passing Standards Recommended by Educators, Parents, and Students from 66 Schools Receiving Materials With Field-Test Performance Data.

Educators			Parents			Students		
n	Mean %Correct Recommended	S.D.	n	Mean %Correct Recommended	S.D.	n	Mean %Correct Recommended	S.D.
594*	73.7	4.3	678*	74.4	4.6	1,194*	75.2	4.2

*Numbers of individuals contributing to school averages.

Table 3. Means and Standard Deviations of Passing Standards Recommended by Educators, Parents, and Students from All 124 Responding Schools.

Educators			Parents			Students		
n	Mean %Correct Recommended	S.D.	n	Mean %Correct Recommended	S.D.	n	Mean %Correct Recommended	S.D.
1,255*	75.6	4.8	1,268*	75.9	4.6	2,206*	75.9	4.3

*Numbers of individuals contributing to school averages.

A comparison of Tables 1 and 2 reveals that the presence of the performance data sheet appears to lower the standards recommended by educators, parents, and students. Educators who received the performance data sheet recommended a passing standard of 73.5%, while their counterparts who had no performance data recommended a passing standard of 77.3%. This difference, when analyzed via a t test, was significant beyond the .01 level ($t = 5.1$). Parents receiving performance data recommended a passing standard of 73.5%, while those parents receiving no performance data recommended a 77.8% standard ($t = 4.1$, $p < .01$). Students who received performance data recommended a 74.9% passing standard, while students who received no performance data recommended a 78.0% passing standard ($t = 2.1$, $p < .05$). Given the consistency of the differences with the three groups, it would surely seem that these particular performance data, in the form presented to participants in this study, did depress the standards recommended by educators, parents, and students.

In summary, sets of recommended performance standards for the MFMT

were secured from 4,729 individuals in Maryland. There were only slight differences among the standards recommended by educators, parents, and students. The knowledge of field-test performance data given to approximately one half of those supplying preferences tended to lower the standards recommended.

MARYLAND FUNCTIONAL WRITING TEST

As mentioned earlier, MSDE also administers the Maryland Functional Writing Test for graduation certification. This instrument consists of two writing prompts, one to elicit an explanatory response and one for a narrative response. The two writing passages will be scored using a four point modified-holistic procedure. Setting a standard for this test has many different problems than those presented by an objectively scored test.

Using an item-based judgmental routine similar to the one used for mathematics requires a rethinking of some procedures. Should judges set a standard looking only at writing prompts rather than student responses? If judges see student writing, should they be given the holistic scores attached to the responses? Should a standard be set on each domain or on a combination of scores from both the explanatory and narrative prompts?

These and many other questions must be answered before actual procedures can begin. However, the need for maximum participation and well-informed judges must be uppermost in discussion.

CONCLUSION

MSDE has not, as yet, chosen a specific standard for the MFMT. This activity will occur in March, 1983. Although the particular standards chosen are very important to all educational personnel involved, the system of standard-setting must also be explained and defended. Livingston & Zieky (1982) ask seven questions about the social and political issues of the standard setting process. To conclude this paper, I would like to answer these questions for Maryland's Functional Testing Program.

1) "Should You Allow Exceptions to Your Decision Rule?"

   The only exceptions allowed by MSDE are based upon "invalid" scores. If some interference prevents adequate measurement of achievement, the score is labeled invalid rather than fail and the student is tested during the next regularly scheduled testing period.

2) "Should You Allow Test-takers Who Fail the Test to Take it Again?"

   Yes. Since certification exams are administered in the ninth grade, retesting is an intimate part of the Maryland system. Diagnostic scores that emerge from the testing are used to guide appropriate instructional assistance to those students who have not passed the examination.

3) "Should Persons Who Have Passed the Test Ever Have to Take It Again?"

No. Maryland has adopted the position "Once a master, always a master."

4) "Should You Establish An Uncertain Category?"

If only one opportunity were given for meeting the certification requirement, an uncertain category might be compelling. Since multiple opportunities are given to non-passers, the need for an uncertain category is diminished.

5) "Should You Use Normative Data in Setting an Absolute Standard?"

Yes. MSDE chose to present "impact data" to a sample of our judges. The consequences of different standards were considered as essential information for well-informed judgments. This information significantly depressed the recommended standards from all judges.

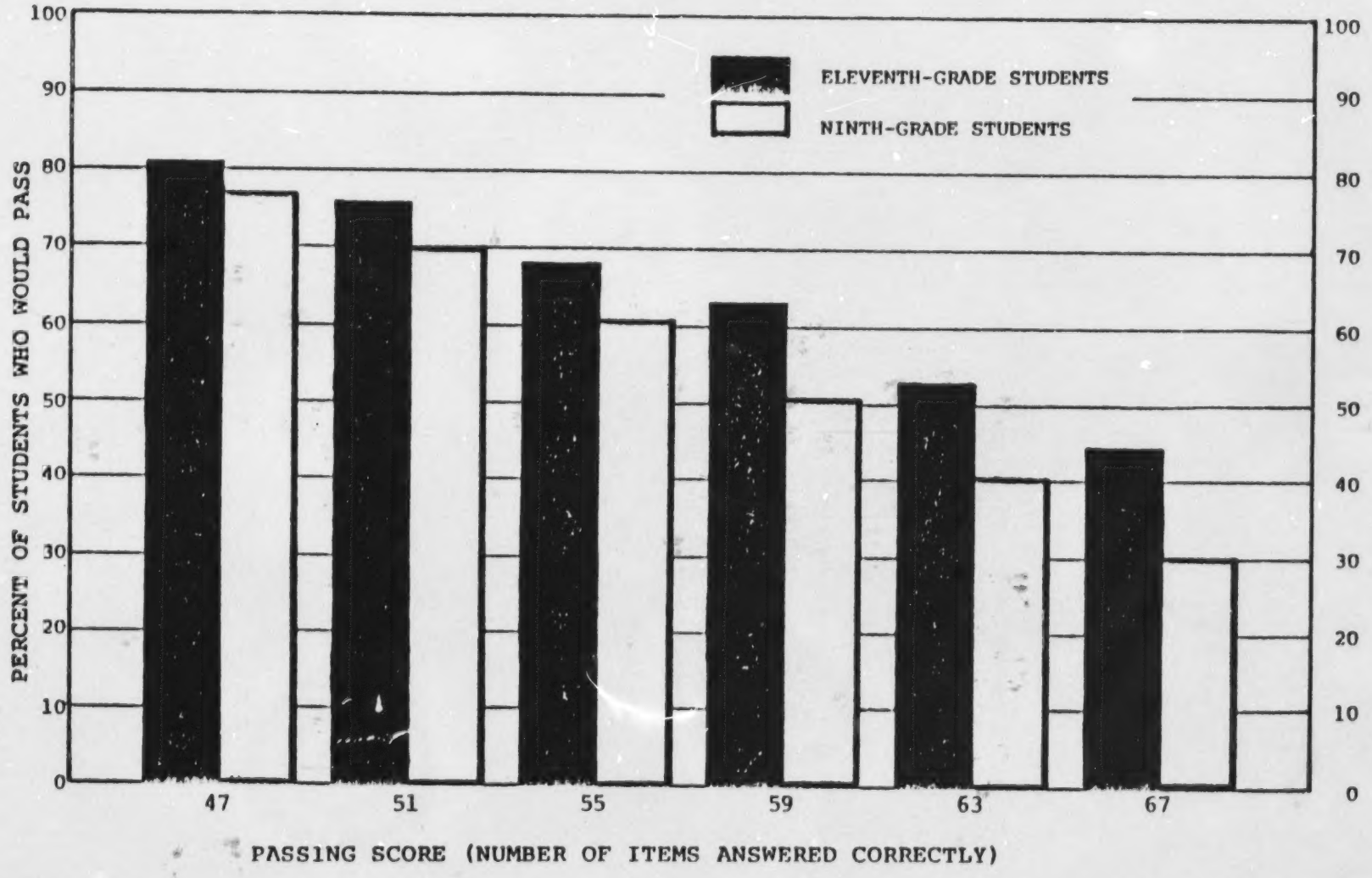
6) "Should You Allow the Standard to Change One Time?"

No. Since MSDE administers alternate forms of all certification exams, the potential for a changing standard is there. To control for variations in difficulty across forms, MSDE equates forms and reports scores in a stable scale score metric. The passing standard for mathematics will be reported as a scale score of 340.

7) "Should You Set Different Passing Scores for Different Groups of Test Takers?"

MSDE has decided not to do this. One aim of the entire Functional Testing Program is to bring clear definitions to the Maryland High School diploma. All students—special—regular and gifted education—must meet the same standards before being awarded a diploma.

PROJECTED PASSING RATES AT DIFFERENT PASSING SCORES



AVERAGE RECOMMENDED PASSING STANDARDS FOR THE MARYLAND FUNCTIONAL MATHEMATICS TEST (MFMT)

These recommendations were made by Citizens, Teachers and Administrators who met for a workshop on Saturday, June 12, 1982 and:

1) Actually took the 1982 form of the MFMT,

2) Received information about the difficulty of each test item on the MFMT when it was given to Maryland ninth graders,

3) Received information about the percentage of ninth graders who would fail the MFMT at various passing scores, and

4) Received information about what the other people in the workshop recommended.

The results were:

Category of Participants	Number of Participants	Average % Correct Recommended
Citizens	28	85%
Teachers	10	78%
Administration	10	84%
Total	48	83.3%

AVERAGE RECOMMENDED PASSING STANDARDS FOR THE MARYLAND FUNCTIONAL MATHEMATICS TEST (MFMT)

These recommendations were made by Educators, Parents and Students from 66 Maryland schools receiving mailout packets containing:

1) A detailed description of the MFMT including sample items, and

2) Information about the percentage of ninth graders who would fail the MFMT at various passing scores:

The results were:

Category of Respondent	Number of Respondents	Average % Correct Recommended
Educators	*594	73.7%
Parents	*678	74.4%
Students	*1,194	75.2%
Total	*2,466	74.6%

*Number of individuals contributing to school averages

FALL 1982 — MARYLAND FUNCTIONAL MATHEMATICS TEST
FREQUENCY DISTRIBUTION OF TOTAL TEST RAW SCORE

Raw Score	Percent Correct	Percent Failing
49	62.8%	27.8%
50	64.1%	29.7%
51	65.4%	31.8%
52	66.7%	34.0%
53	67.9%	36.0%
54	69.2%	38.3%
55	70.5%	40.5%
56	71.7%	42.7%
57	73.0%	45.1%
58	74.3%	47.5%
59	75.6%	50.0%
60	76.9%	52.6%
61	78.2%	55.2%
62	79.5%	57.7%
63	80.8%	60.4%
64	82.1%	63.0%
65	83.3%	65.7%
66	84.6%	68.4%
67	85.9%	71.1%
68	87.2%	73.8%
69	88.5%	76.3%
70	89.7%	79.1%
71	91.0%	81.6%

BIBLIOGRAPHY

Declared Competencies Index, Maryland State Department of Education, Baltimore, Maryland, 1980.

Glass, G.V. Standards and Criteria. Journal of Educational Measurement, 1978, 15 (4), 237-261.

Hambleton, R.K. Test Score Validity and Standard-Setting Methods. In R.A. Berk (Ed.), Criterion-Referenced Measurement: The State of the Art, Baltimore: The Johns Hopkins University Press, 1980.

Jaeger, R.M. An Interactive Structured Judgement Process for Establishing Standards on Competency Tests: Theory and Application. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, April 1981.

Livingston, S.A., & Zieky, M.J. Passing Scores. Princeton, N.J.: Educational Testing Service, 1982.

Popham, W.J. Modern Educational Measurement, Chapter 16, "Setting Standards," Prentice-Hall, Inc., Englewood Cliffs, N.J.: 1981, 371-399.